Principal Component Analysis Based Face Recognition System Using Fuzzy C-means Clustering Classifier

Vivek Banerjee
vivekbanerjee6687@gmail.com

Prof. Prashant Jain
Jabalpur Engineering College

Abstract — Face detection is to find any face in a given image. Face recognition is a two-dimension problem used for detecting faces. The information contained in a face can be analyzed automatically by this system like identity, gender, expression, age, race and pose. Normally face detection is done for a single image but it can also be extended for video stream. As the face images are normally upright, they can be described by a small set of 2-D characteristics views. Here the face images are projected to a feature space or face space to encode the variation between the known face images. In this paper, PCA is used for dimension reduction, the projected feature space is formed using fuzzy c-means clustering algorithm. The above process can be used to recognize a new face in an unsupervised manner. It takes into consideration not only the face extraction but also the mathematical calculations which enable us to bring the image into a simple and technical form. It can also be implemented in real-time using data acquisition hardware and software interface with the face recognition systems. Fuzzy sets can efficiently manage the vagueness and ambiguity of face images degraded by poor illumination component.

Keywords — Face Recognition, Principal Component Analysis, Fuzzy C-means Clustering, Eigen Face, Euclidian Distance.

I. INTRODUCTION

Face recognition plays an important role in many applications such as building/store access control, suspect identification and surveillance [1–9]. Over the past 30 years many different face recognition techniques have been proposed, motivated by the increased number of real-world applications requiring recognition of human faces. There are several problems that make automatic face recognition a very difficult task. The face image of a person input to a face recognition system is not usually taken under similar conditions as the face image of the same person in the database. Therefore, it is important that the automatic face recognition system be able to cope with numerous variations of images of the same face. The image variations are mostly due to changes in the following parameters: pose, illumination, expression, age, disguise, facial hair, glasses, and background [6–9]. Information and Communication Technologies (ICT) are increasingly entering in all aspects of our life and in all sectors, opening a world of unprecedented scenario where people interact with electronic devices, embedded in environments that are sensitive and responsive to the presence of users. Image analysis is a process of discovering, identifying, and understanding patterns that are relevant to the performance of an image-based task. Face recognition has recently received significant attention. It plays an important role in many application areas, such as human-machine interaction, authentication and surveillance.

However, the wide-range variations of human face, due to pose, illumination, and expression, result in a highly complex distribution and deteriorate the recognition performance.

In addition, the problem of machine recognition of human faces continues to attract researchers from disciplines such as image processing, pattern recognition, neural networks, computer vision, computer graphics, and psychology. In identification problems, the input to the system is an unknown face, and the system reports back the determined identity from a database of known individuals, whereas in verification problems, the system needs to confirm or reject the claimed identity of the input face. The solution to the problem involves segmentation of faces (face detection) from cluttered scenes, feature extraction from the face regions, recognition or verification. Robust and reliable face representation is crucial for the effective performance of face recognition system and still a challenging problem. Feature extraction is realized through some linear or nonlinear transform of the data with subsequent feature selection for reducing the dimensionality of facial image so that the extracted feature is as representative as possible.

In this paper, we use PCA for dimension reduction, and a fuzzy c-means clustering, PCA Eigen-face method and Euclidian distance classifier for feature extraction and face recognition.
II. REVIEW OF LITERATURES

Over the periods of time, there were various techniques developed for machine learning and perception. They were categories basically into four techniques namely,

   i) Template matching,
   ii) Prototype matching,
   iii) Distinctive feature comparison and
   iv) Computational techniques.

Usually non-human pattern recognition systems are based on templates. A template system [17] works well for computer that are provided with a standardized set of numbers. But the templates matching are totally inadequate for complex pattern recognition, because it is not flexible. It was used for recognition of only letters and numbers. Here an infinite numbers of templates were necessary to recognize all possible variations, found among numbers and letters. This process could not accommodate the variations of rotation and scaling. Secondly, this approach cannot handle complexities of human visual processing. Therefore, there was a need for more flexible system rather than matching a pattern against a specific template.

The second approach i.e. prototype matching is a flexible version of template matching. Here a set of abstracts and idealized patterns are stored in memory. Matching does not need to be exact, rather variations are allowed. If the match is closed enough, the stimulus is recognized. If the match is inadequate, the stimulus is compared with other prototypes, until a perfect match is located. It includes certain characteristics of facial features. The prototypes are stored in memory not as templates but like features such as hair style, fore head, presence of absence of glass, etc. This method helps to recognize the variety of different representation of the same shape, variety of orientation of the shapes and fragmented view of shapes. It consists of the features common to all or most of the instance of pattern, such as a round heads, high forehead, long eye, high mouth etc.. But the demerit of this technique is that it is not full proof. Research are still going on, to find the exact matching algorithm. Distinctive feature approach states that discrimination are made on the basis of small number of characteristics. These characteristics differentiate one entity from another are called distinctive features. A list of feature components for each entity are stored. Here the pattern recognition involves detecting specific important part of stimulus in contrast to the other matching models. There are many literatures available pertaining to this approach. We will be discussing few of them below.

Computation Approach aims at rapid and accurate recognition of three dimensional objects. The use of computers to simulate this perceptual processes is known as machine vision. There are various machine learning approaches for the 3D object recognition [29]. Our model is based on prototype matching and distinctive feature approach. We will elaborate it in next section. Before that let us briefly analyze few of the existing methods.

Brunelli and Poggio in their paper [13], implemented two algorithms. One is based on the computation of a set of geometrical features, such as nose width and length, mouth position and chin shape and the other is based on the gray-level template matching. They have reported that the face recognition based on geometrical features gives around 90% correct recognition whereas template matching gives perfect recognition. But template matching is very time consuming and cannot handle large database of facial image. On the other hand, the feature based face recognition can handle large database of faces. But gives a result which is less than 100%. Krisnamurthy and Ranganath have reported a feature based face recognition system by using the Eigen-spaces and Wavelet approaches [23], which gives comparatively good result. Chaddha, et. al. proposed a Neural Network based technique [15], for recognition of human faces, using the outline profile of the end view of the human face. This pattern yields a set of good discriminate features for identification. Here nose point, chin points, forehead point, bridge points, nose bottom points, throat points, upper lip point, mouth of center lip point, chin curve point or lip bottom point, brow points are extracted and they are trained by using a back propagation Neural Networks. As a new face is given as the input of the network, it is able to recognize the face. The
accuracy of the system is not good, as it was taking only the side face features. Sinha P. proposed an adaptive Neural network based approach [28], by which the pair of images to be compared should be presented to the Neural network as source (input) and target images. The Neural network learns about the symmetry between the pair of images by analyzing the examples of associated feature pairs belonging to the source and target images. From such pairs of associated features, the Neural Network searches out proper locations of target features from the set of ambiguous target features by fuzzy analysis during its learning. If any of target features searched out by the Neural network, lies outside the prescribed zone, the training of the Neural network is unsuccessful. In case of successful training, the neural network gets adapted with appropriate symmetry relation between the pair of images. When the source image is input to the trained Neural network, it gives an processed source image as output which is later superimposed in target image and identity is established. Zhang and Fulcher reported a face recognition system using Artificial Neural Network Group-based Adaptive Tolerance (GAT) Trees [34]. It is a hierarchical classification, maps on to binary tree structures where each leaf node corresponds to a separate category of faces. Decisions are made in descending down the tree, through each intermediate node, as to whether the current input sample belongs to a specific subclass or not. Only an Nlevels are required in practice to discriminate between a large (2N) categories. Robert and Ritter in their paper [26], propose an Artificial neural network method for human face recognition. They used three neural networks of local linear maps types, which enables a machine to identify the head orientation of a user by learning from examples. One network is used for colour segmentation, the second for localization of the face and the third for the final recognition of the head orientation. This technique could not gain importance, because it suffers from mapping problems.

Goudail F. and et. al. in their paper [17], investigated the face recognition method based on the computation of Autocorrelation coefficients. Auto correlation coefficients are computationally inexpensive, inherently shift-invariant and robust against changes in facial expression. But it can be further extended by segmentation of module based on template matching. T. Kondo and H. Yan reported a feature based face recognition using Cross correlation [22]. Wiskott, et. al. proposed a technique [31], where faces are represented by labelled graphs, based on Gabor Wavelet transform. Image graphs are extracted by an elastic bunch graph matching process and can be compared by a simple similarity function. It is a general method, used for recognising members of a known class of objects. But it is no way specialised to faces. Later Yeal Adini [9] improved the method using 2D Garber like function, which can recognise face with respect to the change in illumination condition. Larry S. Davis went a little further and proposed a method [32] to recognise to human facial expression by long image sequence optical flow. Narendra Ahuja [10] proposed a transformation technique to extract image regions at all geometric and photometric scales. It is intended as a solution to the problem of multiscale, integrated edge and region detection or low level image segmentation. Lanitis and et. al. have reported in their paper [24], recognition of human faces using shape and gray level information. In their approach, they have represented the face by using a small number of parameters, which can be used to code the overall appearance of faces for classification. The approach control both inter class and intra class variations. Discriminant analysis techniques are employed to enhance the effect. This method reported a good result w.r.t. variation of viewpoint, illumination and facial expression.

Now, broadly speaking we can divide the human face recognition process into two categories, i.e. from the geometrical features (such as nose width, mouth position and chin shape) or from gray level template matching. In geometric feature-based matching, a face can be recognized even when the details of the individual features are no longer restored (such as eyes, nose, mouth). The remaining information is, in a sense, purely geometrical and presents, what is left at a very course resolution. In template matching, the image, which is represented
as a bidirectional array of intensity values, is compared using a suitable metric (typically the Euclidean distance) with a single template representing the whole face. Several full templates may be required for the recognition from different viewpoints. On the other hand, if for a single viewpoint, multiple template is considered, then a face is stored as a set of distinctive smaller templates. Here a single template is used together with a qualitative prior model of how a generic face transforms under a change of viewpoint. The deformation model is then heuristically built into the metric, used by the matching measure. This underlying idea is popularly known as elastic templates matching.

In geometric feature based Matching, the extracted features must be somehow normalized in order to be independent of position, scale, and rotation of the face in the mage plane. Translation dependency can be eliminated once the origin of co-ordinates is set to a point that can be detected with good accuracy in each image. To achieves scale and rotation invariance by setting the inter-ocular distance and the direction of the eye-to-eye axis. The normalization rescales the template and image energy distribution so that their average and variances match. To make collection more robust against illumination gradients, each image has to pre-normalized by dividing each pixel by the average intensity over a suitable large neighborhood. As Correlation is computationally expensive therefore, it is performed starting from the lowest resolution level progressively reducing the area of computation from level to level by keeping only a progressively smaller area.

III. OVERVIEW OF SYSTEM

Figure 1 gives an overview of our propose face recognition system. The input consist of image from testing database and there are different images of that person with different expressions and lighting conditions. The system consists of four major blocks: PCA for dimension reduction, fuzzy clustering section for computation of membership function, PCA eigen face for feature extraction, and Euclidian distance classifier. Section 5 introduces the PCA and FCM approach for face recognition, and we evaluate our proposed approach and the paper ends with a conclusion.

IV. DESCRIPTION ON PROPOSED METHODOLOGY

**Principal Component Analysis**

Principal component analysis transforms a set of data obtained from possibly correlated variables into a set of values of uncorrelated variables called principal components. The number of components can be less than or equal to the number of original variables. The first principal component has the highest possible variance, and each of the succeeding component has the highest possible variance under the restriction that it has to be orthogonal to the previous component. We want to
find the principal components, in this case eigenvectors of the covariance matrix of facial images.

The first thing we need to do is to form a training data set. 2D image $I_i$ can be represented as a 1D vector by concatenating rows [2]. Image is transformed into a vector of length $N = mn$.

Let $M$ such vectors $x_i (i = 1, 2, \ldots, M)$ of length $N$ form a matrix of learning images, $X$. To ensure that the first principal component describes the direction of maximum variance, it is necessary to center the matrix. First we determine the vector of mean values $Ψ$, and then subtract that vector from each image vector.

Averaged vectors are arranged to form a new training matrix (size $N \times M$); $A = (φ_1 \ φ_2 \ldots \ φ_M)$.

The next step is to calculate the covariance matrix $C$, and find its eigenvectors $e_i$ and eigenvalues $λ_i$:

$$C = \frac{1}{M} \sum_{i=1}^{M} φ_i φ_i^T = AA^T,$$

$$Ce_i = λ_i e_i.$$

Covariance matrix $C$ has dimensions $N \times N$. From that we get $N$ eigenvalues and eigenvectors. For an image size of 128x128, we would have to calculate the matrix of dimensions 16.384x16.384 and find 16.384 eigenvectors. It is not very effective since we do not need most of these vectors. Rank of covariance matrix is limited by the number of images in learning set — if we have $M$ images, we will have $M-1$ eigenvectors corresponding to non-zero eigenvalues. One of the theorems in linear algebra states that the eigenvectors $e_i$ and eigenvalues $λ_i$ can be obtained by finding eigenvectors and eigenvalues of matrix $C1 = A^TA$ (dimensions $M \times M$) [3]. If $v_i$ and $μ_i$ are eigenvectors and eigenvalues of matrix $A^TA$, then:

$$A^TAv_i = μ_i v_i$$

Multiplying both sides of above equation with $A$ from the left, we get:

$$AA^TAv_i = Aμ_i v_i,$$

$$AA^T(Av_i) = μ_i (Av_i),$$

$$C(Av_i) = μ_i (Av_i).$$

Comparing equations, we can conclude that the first $M-1$ eigenvectors $e_i$ and eigenvalues $λ_i$ of matrix $C$ are given by $Av_i$ and $μ_i$, respectively. Eigenvector associated with the highest eigenvalue reflects the highest variance, and the one associated with the lowest eigenvalue, the smallest variance. Eigenvalues decrease exponentially so that about 90% of the total variance is contained in the first 5% to 10% eigenvectors [3]. Therefore, the vectors should be sorted by eigenvalues so that the first vector corresponds to the highest eigenvalue. These vectors are then normalized. They form the new matrix $E$ so that each vector $e_i$ is a column vector.

The dimensions of this matrix are $N \times D$, where $D$ represents the desired number of eigenvectors. It is used for projection of data matrix $A$ and calculation of $y_i$ vectors of matrix $Y = (y_1 \ldots y_M)$:

$$Y = E^TA$$

Each original image can be reconstructed by adding mean image $Ψ$ to the weighted summation of all vectors $e_i$. The last step is the recognition of faces. Image of the person we want to find in training set is transformed into a vector $P$, reduced by the mean value $Ψ$ and projected with a matrix of eigenvectors (eigenfaces):

$$Ω=E^T(P-Ψ).$$

Classification is done by determining the distance, $ε$, between $Ω$ and each vector $yi$ of matrix $Y$. The most common is the Euclidean distance, but other measures may be used. This paper presents the results for the Euclidean distance. If $A$ and $B$ are two vectors of length $D$, the distance between them is determined as follows [4]:
Euclidean distance:

\[ d(A, B) = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2} = \|A - B\|. \]

If the minimum distance between test face and training faces is higher than a threshold \( \theta \), the test face is considered to be unknown, otherwise it is known and belongs to the person \( s = \text{argmin}_i [\varepsilon_i] \).

**Fuzzy clustering** is a class of algorithms for cluster analysis in which the allocation of data points to clusters is not "hard" (all-or-nothing) but "fuzzy" in the same sense as fuzzy logic. Data clustering is the process of dividing data elements into classes or clusters so that items in the same class are as similar as possible, and items in different classes are as dissimilar as possible. Depending on the nature of the data and the purpose for which clustering is being used, different measures of similarity may be used to place items into classes, where the similarity measure controls how the clusters are formed. Some examples of measures that can be used as in clustering include distance, connectivity, and intensity.

In hard clustering, data is divided into distinct clusters, where each data element belongs to exactly one cluster. In fuzzy clustering (also referred to as soft clustering), data elements can belong to more than one cluster, and associated with each element is a set of membership levels. These indicate the strength of the association between that data element and a particular cluster. Fuzzy clustering is a process of assigning these membership levels, and then using them to assign data elements to one or more clusters.

In fuzzy c-means clustering, each point has a degree of belonging to clusters, as in fuzzy logic, rather than belonging completely to just one cluster. Thus, points on the edge of a cluster, may be in the cluster to a lesser degree than points in the center of cluster. An overview and comparison of different fuzzy clustering algorithms is available.

Any point \( x \) has a set of coefficients giving the degree of being in the \( k \)th cluster \( w_k(x) \). With fuzzy c-means, the centroid of a cluster is the mean of all points, weighted by their degree of belonging to the cluster:

\[ c_k = \frac{\sum_x w_k(x) x}{\sum_x w_k(x)}. \]

The degree of belonging, \( w_k(x) \), is related inversely to the distance from \( x \) to the cluster center as calculated on the previous pass. It also depends on a parameter \( m \) that controls how much weight is given to the closest center. The fuzzy c-means algorithm is very similar to the k-means algorithm:

- Choose a number of clusters.
- Assign randomly to each point coefficients for being in the clusters.
- Repeat until the algorithm has converged (that is, the coefficients' change between two iterations is no more than \( \varepsilon \), the given sensitivity threshold)
- Compute the centroid for each cluster, using the formula above.
- For each point, compute its coefficients of being in the clusters, using the formula above.

The algorithm minimizes intra-cluster variance as well, but has the same problems as k-means; the minimum is a local minimum, and the results depend on the initial choice of weights.

Using a mixture of Gaussians along with the expectation-maximization algorithm is a more statistically formalized method which includes some of these ideas: partial membership in classes. Another algorithm closely related to Fuzzy C-Means is Soft K-means. Fuzzy c-means has been a very important tool for image processing in clustering objects in an image. In the 70's, mathematicians introduced the spatial term into the FCM algorithm to improve the accuracy of clustering under noise.

- More sophisticated usage of class assignment of patterns (faces).
- Classification results affect the within-class and between-class scatter matrices.

**Algorithm**

*Given set of feature vectors transformed by the PCA, \( X = \{x_1, x_2, \ldots, x_N\}\).*

**Partition matrix**

\( U = [\mu_{ij}] \) for \( i = 1, 2, \ldots, c \) and \( j = 1, 2, \ldots, N \)

Which satisfies, \( \sum_{i=1}^{c} \mu_{ij} = 1 \)

\[ 0 < \sum_{j=1}^{N} \mu_{ij} < N \]
The Computations of Membership Degrees

- Compute the Euclidean distance matrix between pairs of feature vectors in the training.
- Set diagonal elements of this matrix to infinity.
- Sort the distance matrix in ascending order.
- Collect the class labels of the patterns located in the closest neighborhood of the pattern.
- Compute the membership grade to class \( i \) for \( j \)th pattern,

\[
\mu_{ij} = \begin{cases} 
0.51 + 0.49(n_{ij} / k) & \text{if } i = \text{same as the label of the } j\text{th pattern} \\
0.49(n_{ij} / k) & \text{if } i \neq \text{same as the label of the } j\text{th pattern}
\end{cases}
\]

Where, \( n_{ij} \) is number of the neighbors of the \( j \)th data that belong to \( i \)th class.

Procedural Steps

- Results of fuzzy C-means clustering classification are used in computations of mean value and scatter covariance matrices.
- Mean vector of each class

\[
m_i = \frac{\sum_{j=1}^{N} \mu_{ij} x_j}{\sum_{j=1}^{N} \mu_{ij}}
\]

- The between class and within class fuzzy scatter matrices are respectively,

\[
S_{11} = \sum_{i=1}^{c} N_i (m_i - m)(m_i - m)^T
\]

\[
S_{12} = \sum_{i=1}^{c} \sum_{x \in c_i} N_i (m_i - m)(m_i - m)^T = \sum_{i=1}^{c} S_{12i}
\]

- The optimal fuzzy projection \( W_{\text{main}} \) and feature vector transformed by fuzzy clustering based method are given by

\[
W_{\text{main}} = \arg \max_w \frac{[W^T, S_{11}W]}{[W^T, S_{11}W]}
\]

\[
V_i = W_{\text{main}}^T x_i = W_{\text{main}}^T E^T (Z_i - Z)
\]

- Fuzzy clustering approach outperform the other methods for the datasets considered.
- Sensitivity variations in illumination and facial expression reduced substantially.
- Fuzzy sets can efficiently manage the vagueness and ambiguity of face images degraded by poor illumination component.
V. EXPERIMENTAL RESULTS

Figure 2: Pictures from the training base.

Figure 3: Shows all eigenvalues. Each eigenvalue corresponds to a single eigenvector and tells us how much images from training bases vary from the mean image in that direction.

Figure 4: An Example of Test images and recognized images from the training database.

VI. CONCLUSION & DISCUSSION

Human face detection and recognition have drawn considerable interest and attention from many researchers for decades. It has several applications such as criminal investigation, authentication in secure system bank teller machines etc. The first step of face recognition process is the face location process or briefly saying eye location step. Accurate detection of eye components will enable the development of an accurate face recognition system.

When face orientation is more than 20% towards either side of normal to the plane, the eye detection fails. In this situation, mouth can be taken as a reference object. A standard detection technique such as normalized correlation template matching is one among the few techniques used for face recognition.

In this paper we have exploited the fuzzy clustering technique for the face recognition process, along with the use of principal component analysis. The experiment shows a result of 98% with the free environment.

When frontal images are used, registered person’s recognition rate is over 98% (in 310 test images, 4 images are false, and 1 image is rejected). Furthermore, non-registered person’s rejected rate of 99% (in 100 test images, 1 image is false) is obtained. The results have demonstrated the promising performance of this algorithm. However, under different perspective variation, the registered person’s recognition is 74.19% (in 310 images, 16 images are false, and 64 images are rejected).
• Fuzzy clustering approach outperform the other methods for the datasets considered.
• Sensitivity variations in illumination and facial expression reduced substantially.
• Fuzzy sets can efficiently manage the vagueness and ambiguity of face images degraded by poor illumination component.

REFERENCES
